Estimation of Non-Linear Expectile Regression Model with Censoring Indicators Missing at Random

Ruiping Hou, Shuanghua Luo*, Junxiang Lu

***Abstract*—The present paper focuses on the expectile regression for a non-linear model with right-censored responses and censoring indicators that are missing at random (MAR). Firstly, a calibration estimator and an imputation estimator are constructed for the non-linear expectile regression model. Secondly, under specific regularity conditions, the asymptotic normality of the estimators is established. The validity of the proposed estimation methods is finally verified through simulation and numerical experiments.**

***Index Terms*—censoring indicators, missing at random, non-linear model, expectile regression, asymptotic normality**

I. Introduction

In the domain of regression analysis, the sensitivity of the least squares regression method to abnormal data is a critical consideration. When the error distribution deviates from the standard normal distribution, the effectiveness of mean regression estimation is significantly weakened. Koenker et al. proposed quantile regression (QR) in 1978, has already become an important statistical method for data analysis. However, it should be emphasized that the loss function associated with QR is not differentiable. Within specific complex modeling frameworks, the direct implementation of quantile regression may induce bias in parameter estimation - a consequence that circumscribes its practical applicability across diverse real-world scenarios. To address these limitations, expectile regression (ER), proposed by Newey et al. [1], adopts the asymmetric sum of squared residuals as its loss function. This approach enables the accurate estimation of heteroskedasticity embedded in mean - based regression frameworks and places heightened emphasis on the tail characteristics of the response variable. Waltrup et al. [2] conducted a comparison between expectile regression and quantile regression, highlighting that the loss function of expectile regression is convex and differentiable - representing a key advantage in the optimization of complex models. Quantile regression is regarded as a generalization of median regression. In contrast, expectile regression is regarded as a generalization of mean regression. For instance, in simple parametric models, Sobotka et al. [3] established the asymptotic normality of expectile regression estimates. Kim et al. [4] studied the asymptotic properties of asymmetric least squares estimates in non-linear models with heteroscedasticity. This line of research has significantly advanced the development of estimation frameworks for expectile regression. In a further contribution, Jiang et al. [5] proposed a penalized asymmetric least squares estimator for the single-index expectile model, aiming to enhance the model's flexibility in high-dimensional scenarios. Gao et al. [6] developed a two-stage premium calculation model. The first stage of this model adopts logistic regression to estimate the probability of at least one claim, while the second stage integrates generalized linear models with parametric expectile regression to refine risk assessment. Litimein et al. [7] explored the non-parametric estimation of functional expectile regression.

The aforementioned studies uniformly assume the availability of complete datasets. However, in the fields of survival analysis and clinical research, constraints from limitations in experimental design and restrictions on observation time often prevent researchers from accurately observing event-specific survival times. For instance, in medical trials, the follow-up duration is usually pre-determined. Patients are enrolled in the cohort in a relatively random manner; some may withdraw early or lose to follow-up before the study concludes, while others may die from causes unrelated to the disease under investigation. Consequently, the collection of survival data is frequently incomplete, resulting in censored or missing data. A significant number of scholars have previously conducted research on data incompleteness. For example, Ji et al. [8] proposed a single-index varying coefficient quantile model where covariates are MAR. Koul et al. [9] introduced expectile regression in a linear model under the condition of randomly right-censored observations. Meanwhile, Pan et al. [10] proposed a weighted expectile regression method to estimate the conditional expectile in the presence of covariates are missing at random. Seipp et al. [11] proposed an extension of expectile regression that incorporates inverse probability weights to address right-censored data. Furthermore, Zhao et al. [12] introduced an improved weighted expectile average estimator based on the covariate balancing propensity score for linear models with response variables missing at random. Zhang et al. [13] developed a novel weighted expectile regression neural network method,

Manuscript received April 8, 2025; revised August 2, 2025.

This work was supported by the Natural Science Foundation of Shaanxi Province (China) (2024JC-YBMS-007; 2024JC-YBMS-037).

Ruiping Hou is a postgraduate student at the School of Science, Xi'an Polytechnic University, Xi'an 710048 China. (e-mail: 18591977851@163.com).

Shuanghua Luo is a professor at the School of Science, Xi'an Polytechnic University, Xi'an 710048 China. (Corresponding author, e-mail: iwantflyluo@163.com).

Junxiang Lu is a professor at the School of Science, Xi'an Polytechnic University, Xi'an 710048 China. (e-mail: jun-xianglu@163.com).

which integrates the inverse probability of censoring weighting technique into the expectile loss function to address random censoring problems. Ciuperca [14] proposed and investigated a random right-censoring model estimated via the expectile method, using the expectile loss function and the adaptive lasso penalty.

The aforementioned papers all assume that the censoring indicators are always observable. In fact, in many practical scenarios, the censoring indicators may not be fully observed because various reasons. For instance, Wang et al. [15] studied least squares regression in a linear model using regression calibration and imputation when some censoring indicators are missing. Shen et al. [16] proposed quantile regression for a partially linear varying-coefficient model using regression calibration and imputation when the responses are right-censored and the censoring indicators missing at random. Wang et al. [17] conducted a study on weighted composite quantile regression in a linear model with right-censored data and censoring indicators missing at random. Zhou et al. [18] developed a varying-coefficient partially non-linear quantile regression model for right-censored response variables with censoring indicators missing at random. To the best of our knowledge, existing research on expectile regression estimation remains limited in this context. Therefore, studying the expectile regression under the scenario of censored data and censoring indicators missing at random holds clear practical significance.

Building upon the research of Seipp et al. [11], this paper extends the scenario of right-censored data to the case where responses are right-censored and censoring indicators missing at random, and then focuses on studying the expectile regression estimation for non-linear models. Firstly, we construct a calibration estimator and an imputation estimator within the framework of the non-linear expectile regression model. Subsequently, under certain assumptions, the asymptotic properties of the parameters estimated by different methods are established. Finally, the effectiveness of the proposed estimation methods is verified through simulation studies and real data analyses.

The rest of this paper is organized as follows. In Section 2, we introduce the expectile regression method in the non-linear model when the data are the right-censored and the censoring indicators are MAR. The asymptotic properties of the main results are established in Section 3 under certain suitable conditions. A simulation study is presented to evaluate the performance of the proposed methods in Section 4. We apply the proposed methods to analyze data from the German breast cancer study group in Section 5.

II. Model and Estimators

In this paper, the following non-linear model will be considered:

$$T_i = f\left(X_i, \beta\right) + \varepsilon_i, i = 1, 2, 3, \cdots, n \quad (1)$$

where T_i denotes the response variable, X_i is a random vector of covariates, $\beta \in \mathbb{R}^p$ represents an unknown parameter, ε_i is the error terms, and $f(\cdot)$ is a known non-linear function.

In Survival Analysis, T_i is usually logarithm of survival time. Let C_i be the censoring variable, and its distribution function being $G(\cdot)$, $Y_i = \min\left(T_i, C_i\right)$, and censoring indicators $\delta_i = I\left(T_i \le C_i\right)$. We suppose that Y_i and C_i are independent conditional under X_i. For simplicity, define a missing indicator ξ_i, if $\xi_i = 1$, when δ_i is observed, and $\xi_i = 0$ when δ_i is missing. Then, we can observe an independent identically distributed (i.i.d.) sample $\left(Y_i, X_i, \xi_i\delta_i, \xi_i, 1 \le i \le n\right)$. We assume T_i and C_i are mutually independent, and δ_i is missing at random (MAR), which means that δ_i and ξ_i are conditionally independent given $\left(Y_i, X_i\right)$, that is

$$P\left(\xi_i = 1 \middle| Y_i, X_i, \delta_i\right) = P\left(\xi_i = 1 \middle| Y_i, X_i\right) := \Delta\left(Y_i, X_i\right).$$

Let $E_T\left(\tau \middle| X\right)$ be the τ conditional expectile of T given X, then $E_T\left(\tau \middle| X\right) = \arg\min_{\alpha} E\left(\rho_\tau\left(T - \alpha\right) \middle| X\right)$, where $\rho_\tau\left(u\right) = u^2\left|\tau - I\left(u < 0\right)\right|$ is an asymmetric squared loss function, $I(\cdot)$ denotes the indicator function, $\tau \in (0,1)$ is the expectile index. The derivative of $\rho_\tau(u)$ is

$$\rho_\tau'\left(u\right) = 2\left|\tau - I\left(u < 0\right)\right|u.$$

Under the presence of censoring, given the conditional independence of Y_i and C_i given X_i, we derive

$$E\left\{\frac{\delta_i}{1 - G\left(Y_i\right)}\rho_\tau\left(Y_i - \alpha\right) \middle| X_i\right\} = E\left\{\rho_\tau\left(T_i - \alpha\right) \middle| X_i\right\}.$$

Thus, it can be concluded that the expectile regression estimator of β can be defined by solving the minimization problem of the following objective function:

$$\sum_{i=1}^{n} \frac{\delta_i}{1 - \hat{G}\left(Y_i\right)}\rho_\tau\left(Y_i - f\left(X_i, \beta\right)\right), \quad (2)$$

where $\hat{G}(\cdot)$ is the Kaplan-Meier estimator [19].

However, since the censoring indicator δ_i is missing at random. By replacing δ_i with its conditional expectation $m\left(Y_i, X_i\right) = E\left(\delta_i \middle| Y_i, X_i\right)$, we have

$$E\left\{\frac{m\left(Y_i, X_i\right)}{1 - G\left(Y_i\right)}\rho_\tau\left(Y_i - \alpha\right) \middle| X_i\right\} = E\left\{\frac{\delta_i}{1 - G\left(Y_i\right)}\rho_\tau\left(Y_i - \alpha\right) \middle| X_i\right\}.$$

Thus, the expectile regression estimator of β can be defined by solving the minimization problem of the following objective function

$$\sum_{i=1}^{n} \frac{m\left(Y_i, X_i\right)}{1 - G\left(Y_i\right)}\rho_\tau\left(Y_i - f\left(X_i, \beta\right)\right). \quad (3)$$

However, we should note that two components in (3) are usually unknown in practice: the conditional expectation function $m(\cdot)$ and the censoring distribution function $G(\cdot)$. Thus, we need to estimate them in advance before implementing the estimator.

One usual way to estimate $m(\cdot)$ is to assume it follows a parametric model $m\left(Y, X\right) = m_0\left(Y, X, \theta\right)$, where $m_0(\cdot, \cdot)$ is a known function and θ is an unknown parameter vector.

Following Wang and Dinse [15], the estimator $\hat{\theta}_n$ of θ can be obtained by maximizing the likelihood function

$$\prod_{i=1}^{n} m_0\left(Y_i, X_i, \theta\right)^{\xi_i \delta_i}\left(1-m_0\left(Y_i, X_i, \theta\right)\right)^{\xi_i\left(1-\delta_i\right)}.$$

The definition of $\hat{G}(\cdot)$ was consistent with the approaches outlined in the works of Li and Wang [20] and Wang and Ng [21]. The Nadaraya-Watson estimator of $u(y)=E(\delta|Y=y)$ defined by $\hat{u}_n(y)=\sum_{i=1}^{n}\delta_i\xi_i K\left(\frac{Y_i-y}{h_n}\right)\Big/\sum_{i=1}^{n}\xi_i K\left(\frac{Y_i-y}{h_n}\right)$ with kernel function $K(\cdot)$ and bandwidth sequence $0<h_n \to 0$ as $n\to\infty$. Thus

$$\hat{G}_n(y)=1-\prod_{i:Y_i\le y}\left(\frac{n-R_i}{n-R_i+1}\right)^{1-\hat{u}_n(Y_i)},$$

where $R_i=\sum_{j=1}^{n} I\left(Y_j \le Y_i\right)$.

Let $\delta_i^{ca}=m_0\left(Y_i, X_i, \theta\right)$. We can estimate δ_i^{ca} and $G\left(Y_i\right)$ by $\hat{\delta}_i^{ca}=m_0\left(Y_i, X_i, \hat{\theta}_n\right)$ and $\hat{G}_n\left(Y_i\right)$, respectively. Then, we propose the expectile regression calibration estimator $\hat{\beta}$ defined as follows

$$\hat{\beta}=\arg\min_{\beta}\sum_{i=1}^{n}\frac{\hat{\delta}_i^{ca}}{1-\hat{G}_n\left(Y_i\right)}\rho_\tau\left(Y_i-f\left(X_i,\beta\right)\right). \tag{4}$$

In addition, imputation is often used in statistical analyses with missing data. Next, the expectile regression estimator of the non-linear model is established based on the imputation method. Let $\delta_i^{im}=\delta_i\xi_i+\left(1-\xi_i\right)m_0\left(Y_i, X_i, \theta\right)$, we can estimate δ_i^{im} by $\hat{\delta}_i^{im}=\delta_i\xi_i+\left(1-\xi_i\right)m_0\left(Y_i, X_i, \hat{\theta}_n\right)$. Then, we propose the expectile regression imputation estimator $\hat{\beta}^I$, defined as follows

$$\hat{\beta}^I=\arg\min_{\beta}\sum_{i=1}^{n}\frac{\hat{\delta}_i^{im}}{1-\hat{G}_n\left(Y_i\right)}\rho_\tau\left(Y_i-f\left(X_i,\beta\right)\right). \tag{5}$$

III. Assumptions And Main Results

A. Main Results

The following notations are needed to state the results.

$H(\cdot)$ is the distribution function of Y, $\bar{H}=1-H$. Define

$a_H=\inf\{t: H(t)=1\}$, and $a_G=\inf\{t: G(t)=1\}$,

$$\nabla m_0(Y,X,\theta)=\left(\frac{\partial m_0(Y,X,\theta)}{\partial\theta_1},\cdots,\frac{\partial m_0(Y,X,\theta)}{\partial\theta_l}\right),$$

$$I(\theta)=E\left\{\frac{\xi\nabla m_0(Y,X,\theta)\left[\nabla m_0(Y,X,\theta)\right]^T}{m_0(Y,X,\theta)\left[1-m_0(Y,X,\theta)\right]}\right\},$$

$$\Sigma=E\left[\frac{\partial f\left(X_i,\beta\right)}{\partial\beta}\frac{\partial^T f\left(X_i,\beta\right)}{\partial\beta}\right],$$

$$\Omega_1=E\left\{\frac{\left(m_0(Y,X,\theta)\right)^2}{\left[1-G(Y)\right]^2}\frac{\partial f(X,\beta)}{\partial\beta}\frac{\partial^T f(X,\beta)}{\partial\beta}\left[\rho_\tau'\left(\varepsilon_i\right)\right]^2\right\},$$

$$\Omega_2=E\left\{\frac{\Delta(Y,X)m_0(Y,X,\theta)\left(1-m_0(Y,X,\theta)\right)}{\left[1-G(Y)\right]^2}* \frac{\partial f(X,\beta)}{\partial\beta}\frac{\partial^T f(X,\beta)}{\partial\beta}\left[\rho_\tau'\left(\varepsilon_i\right)\right]^2\right\},$$

$$W=-\sum_{i=1}^{n}\frac{\hat{\delta}_i^{ca}}{1-\hat{G}_n\left(Y_i\right)}\frac{\frac{\partial f\left(X_i,\beta\right)}{\partial\beta}}{\sqrt{n}}\rho_\tau'\left(\varepsilon_i\right),$$

$$L_{in}(\gamma)=\frac{\hat{\delta}_i^{ca}}{1-\hat{G}_n\left(Y_i\right)}\left[\rho_\tau\left(\varepsilon_i-\frac{\frac{\partial f\left(X_i,\beta\right)}{\partial\beta}\gamma}{\sqrt{n}}\right)-\rho_\tau\left(\varepsilon_i\right)+\frac{\frac{\partial f\left(X_i,\beta\right)}{\partial\beta}\gamma}{\sqrt{n}}\rho_\tau'\left(\varepsilon_i\right)\right].$$

To formulate the main results, it is necessary to impose the following regularity conditions:

A1: $F(\cdot)$ is common cumulative distribution function of ε_i, the errors ε_i are independent and identically distributed. Moreover, $E\left(\varepsilon_i\right)=0$, $E\left(\varepsilon_i^2\left|X_i\right.\right)<\infty$ and $E\left[2(1-\tau)\varepsilon_i I\left(\varepsilon_i\le 0\right)+2\tau\varepsilon_i I\left(\varepsilon_i>0\right)\right]=E\left[\rho_\tau'\left(\varepsilon_i\right)\right]=0$.

A2: $K(\cdot)$ is a kernel function of order 1 with bounded support, satisfies $\int_{-1}^{1}K(u)du=1$, $\int_{-1}^{1}uK(u)du=0$.

A3: The bandwidth h_n satisfies the following asymptotic conditions as $n\to\infty$, $nh_n\to\infty$, $nh_n{}^2\to 0$.

A4: The matrix Σ, Ω_1 and Ω_2 are positive definite.

A5: $\left(Y_i, X_i\right), i=1,2,3,\cdots,n$ are an independent identically distributed (i.i.d) random vector.

A6: $\nabla m_0(Y,X,\theta)$ is continuous at θ. The function $u(\cdot)$ has bound derivatives of order 1, and $I(\theta)$ is positive definite matrix.

A7: $a_H<a_G$, $G(\cdot)$ is continuous. For $t<a_H$, $H(\cdot)$ is continuous.

A8: For any i, there is a compact set E, such that $X_i\in E\subset\mathbb{R}^p$.

A9: $E\left(\|X\|^2\right)<\infty$ and $E\left\{\frac{\left[1-G(Y)\right]^2}{\bar{H}^2(Y)}\right\}<\infty$.

Remark1: In accordance with the findings outlined in the proof of Theorem 2.1 in the article by Li and Wang [20], we derive the following conclusions:

(a).

$$\hat{\theta}_n-\theta=I^{-1}(\theta)\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_i\left[\delta_i-m_0(Y_i,X_i,\theta)\right]\nabla m_0(Y_i,X_i,\theta)}{m_0(Y_i,X_i,\theta)\left[1-m_0(Y_i,X_i,\theta)\right]}$$

$+o_p(1)$,

where $\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_i\left[\delta_i-m_0(Y_i,X_i,\theta)\right]\nabla m_0(Y_i,X_i,\theta)}{m_0(Y_i,X_i,\theta)\left[1-m_0(Y_i,X_i,\theta)\right]}=o_p(1)$.

(b). by (a) and conditions A6,

$$\max_i\left|m_0(Y_i,X_i,\hat{\theta}_n)-m_0(Y_i,X_i,\theta)\right|=o_p(1).$$

Theorem 1: (Asymptotic Normality) Under conditions A1-A9, let β_0 is the true value of parameter β. We have

$$\sqrt{n}\left(\hat{\beta}-\beta_0\right)\xrightarrow{D}N\left(0,\frac{1}{4g^2(\tau)}\Sigma^{-1}\Omega_1\Sigma^{-1}\right),$$

$$\sqrt{n}\left(\hat{\beta}^I-\beta_0\right)\xrightarrow{D}N\left(0,\frac{1}{4g^2(\tau)}\Sigma^{-1}\left(\Omega_1+\Omega_2\right)\Sigma^{-1}\right),$$

where $g(\tau)=(1-\tau)F(0)+\tau(1-F(0))$.

B. Proof of Main Result

To prove the Theorem, we introduce a lemma.

Lemma 1: When all the assumptions are true,

$$\hat{G}_n(y)-G(y)=\frac{1-G(y)}{n}\sum_{j=1}^{n}\psi(Y_j,\delta_j,\xi_j;y)+o_p(n^{-1/2}),$$

where

$$\psi(Y_j,\delta_j,\xi_j;y)=\frac{\left[\xi_j-\pi(Y_j)\right]\left[\delta_j-\mu(Y_j)\right]}{\pi(Y_j)\left[1-H(Y_j)\right]}I(Y_j\le y)$$

, and

$$+\frac{I(Y_j\le y,\delta_j=0)}{1-H(Y_j)}+\int_0^{Y_j\wedge y}\frac{dH_0(s)}{\left[1-\left[H(s)\right]^2\right]}$$

$H_0(t)=P(Y_j>y,\delta_j=0)$.

Lemma 1 is proven in Li and Wang [20] . Moreover, it is also proved that $E\left[\psi(Y_j,\delta_j,\xi_j;Y_i)\middle|Y_i\right]=0$, and

$E\left[\psi^2(Y_j,\delta_j,\xi_j;Y_i)\middle|Y_i\right]\le\frac{\left[1-G(Y_i)\right]^2}{\bar{H}^2(Y_i)}$, for $i\ne j$.

Proof of Theorem 1: We can only prove the result about $\hat{\beta}$, the result of $\hat{\beta}^I$ can be prove similarly. To prove the asymptotic normality of $\sqrt{n}\left(\hat{\beta}-\beta_0\right)$, we can estimate by minimizing the following objective function:

$$\sum_{i=1}^{n}\frac{\hat{\delta}_i^{c\alpha}}{1-\hat{G}_n(Y_i)}\rho_\tau\left\{\left(y_i-f(X_i,\hat{\beta})\right)-\left(y_i-f(X_i,\beta_0)\right)\right\}.$$

Let $\varepsilon_i=y_i-f(X_i,\beta_0)$, $\gamma=\sqrt{n}\left(\hat{\beta}-\beta_0\right)$, Then, it is easy to see that γ is minimized, we have

$$y_i-f(X_i,\hat{\beta})=y_i-f(X_i,\beta_0)-\left[\left(f(X_i,\hat{\beta})-f(X_i,\beta_0)\right)\right]$$

$$=\varepsilon_i-\frac{\eta+\frac{\partial f(X_i,\beta)}{\partial\beta}\gamma}{\sqrt{n}}.$$

We can about the following objective function:

$$G_n(\gamma)$$

$$=E(G_n(\gamma))-E(G_n(\gamma))+G_n(\gamma)$$

$$=E(G_n(\gamma))-E\left(\sum_{i=1}^{n}\frac{\hat{\delta}_i^{c\alpha}}{1-\hat{G}_n(Y_i)}\left\{\left(\rho_\tau\left(\varepsilon_i-\frac{\frac{\partial f(X_i,\beta)}{\partial\beta}\gamma}{\sqrt{n}}\right)-\rho_\tau(\varepsilon_i)\right)+\frac{\frac{\partial f(X_i,\beta)}{\partial\beta}\gamma}{\sqrt{n}}\rho_\tau'(\varepsilon_i)\right\}\right.$$

$$\left.-\sum_{i=1}^{n}\frac{\hat{\delta}_i^{c\alpha}}{1-\hat{G}_n(Y_i)}\frac{\frac{\partial f(X_i,\beta)}{\partial\beta}\gamma}{\sqrt{n}}\rho_\tau'(\varepsilon_i)\right)$$

$$+\sum_{i=1}^{n}\frac{\hat{\delta}_i^{c\alpha}}{1-\hat{G}_n(Y_i)}\left\{\left(\rho_\tau\left(\varepsilon_i-\frac{\frac{\partial f(X_i,\beta)}{\partial\beta}\gamma}{\sqrt{n}}\right)-\rho_\tau(\varepsilon_i)\right)\right.$$

$$\left.+\frac{\frac{\partial f(X_i,\beta)}{\partial\beta}\gamma}{\sqrt{n}}\rho_\tau'(\varepsilon_i)\right\}-\sum_{i=1}^{n}\frac{\hat{\delta}_i^{c\alpha}}{1-\hat{G}_n(Y_i)}\frac{\frac{\partial f(X_i,\beta)}{\partial\beta}\gamma}{\sqrt{n}}\rho_\tau'(\varepsilon_i)\Bigg\}$$

$$=E(G_n(\gamma))-E\left(\sum_{i=1}^{n}L_{in}(\gamma)+W^T\gamma\right)+\left(\sum_{i=1}^{n}L_{in}(\gamma)+W^T\gamma\right)$$

$$=E(G_n(\gamma))+\sum_{i=1}^{n}L_{in}(\gamma)-E\left(\sum_{i=1}^{n}L_{in}(\gamma)\right)+W^T\gamma-E\left(W^T\gamma\right).$$

By the condition A1, we have

$$E\left(W^T\gamma\right)=E\left(-\sum_{i=1}^{n}\frac{\hat{\delta}_i^{c\alpha}}{1-\hat{G}_n(Y_i)}\frac{\frac{\partial f(X_i,\beta)}{\partial\beta}\gamma}{\sqrt{n}}\rho_\tau'(\varepsilon_i)\right)=0,$$

$$VAR(W)=E\left(-\sum_{i=1}^{n}\frac{\hat{\delta}_i^{c\alpha}}{1-\hat{G}_n(Y_i)}\frac{\frac{\partial f(X_i,\beta)}{\partial\beta}}{\sqrt{n}}\rho_\tau'(\varepsilon_i)\right)^2$$

$$=E\left(\frac{\left(m_0(Y,X,\theta)\right)^2}{\left(1-\hat{G}(Y)\right)^2}\frac{\partial f(X,\beta)}{\partial\beta}\frac{\partial f^T(X,\beta)}{\partial\beta}\left[\rho_\tau'(\varepsilon_i)\right]^2\right)$$

$$+o_p(1)$$

$$=\Omega_1.$$

According to Lindeberg-Feller central limit theorem, and condition A5, we have

$$W\xrightarrow{D}N(0,\Omega_1),\tag{6}$$

Now, we calculate $E(G_n(\gamma))$,

$$
\begin{aligned}
&E\left(G_n(\gamma)\right)\\
&=\mathrm{E}\left(\sum_{i=1}^{n}\frac{m_0\left(Y_i,X_i,\theta\right)}{1-G_n\left(Y_i\right)}\left(\rho_\tau\left(\varepsilon_i-\frac{\frac{\partial f\left(X_i,\beta\right)}{\partial\beta}\gamma}{\sqrt{n}}\right)-\rho_\tau\left(\varepsilon_i\right)\right)\right)\\
&+E\left(\sum_{i=1}^{n}\frac{m_0\left(Y_i,X_i,\hat{\theta}_n\right)-m_0\left(Y_i,X_i,\theta\right)}{1-G_n\left(Y_i\right)}*\right.\\
&\left.\left(\rho_\tau\left(\varepsilon_i-\frac{\frac{\partial f\left(X_i,\beta\right)}{\partial\beta}\gamma}{\sqrt{n}}\right)-\rho_\tau\left(\varepsilon_i\right)\right)\right)\\
&+E\left(\sum_{i=1}^{n}\frac{m_0\left(Y_i,X_i,\hat{\theta}_n\right)\left[\hat{G}_n\left(Y_i\right)-G\left(Y_i\right)\right]}{\left[1-G\left(Y_i\right)\right]\left[1-\hat{G}_n\left(Y_i\right)\right]}*\right.\\
&\left.\left(\rho_\tau\left(\varepsilon_i-\frac{\frac{\partial f\left(X_i,\beta\right)}{\partial\beta}\gamma}{\sqrt{n}}\right)-\rho_\tau\left(\varepsilon_i\right)\right)\right)\\
&=E\left(G_{1n}(\gamma)\right)+E\left(G_{2n}(\gamma)\right)+E\left(G_{3n}(\gamma)\right).
\end{aligned}
$$

By the condition A7, we have $\frac{m_0\left(Y_i,X_i,\theta\right)}{1-G\left(Y_i\right)}\le C$. Therefore

$$
\begin{aligned}
&E\left(G_{1n}(\gamma)\right)\\
&\le \mathrm{E}\left(\sum_{i=1}^{n}C\left(\rho_\tau\left(\varepsilon_i-\frac{\frac{\partial f\left(X_i,\beta\right)}{\partial\beta}\gamma}{\sqrt{n}}\right)-\rho_\tau\left(\varepsilon_i\right)\right)\right)\\
&\triangleq \mathrm{E}\left(\sum_{i=1}^{n}\left(\rho_\tau\left(\varepsilon_i-\frac{\frac{\partial f\left(X_i,\beta\right)}{\partial\beta}\gamma}{\sqrt{n}}\right)-\rho_\tau\left(\varepsilon_i\right)\right)\right)\\
&\triangleq E\left(\tilde{G}_{1n}(\gamma)\right).
\end{aligned}
$$

By the condition A1, we can obtain the function $M(t)=E\left(\rho_\tau\left(\varepsilon_i-t\right)-\rho_\tau\left(\varepsilon_i\right)\right)$ has a unique minimizer at zero, and its Taylor expansion at the origin has the form $M(t)=g(\tau)t^2+o\left(t^2\right)$ given that $M(0)=0$, and

$$
\begin{aligned}
\nabla_t M(0)&=-E\left[2(1-\tau)\varepsilon_i I\left(\varepsilon_i\le 0\right)+2\tau\varepsilon_i I\left(\varepsilon_i>0\right)\right]\\
&=-E\left[\rho_\tau'\left(\varepsilon_i\right)\right]=0
\end{aligned};
$$

$$
\begin{aligned}
\nabla_t^2 M(0)&=2(1-\tau)E\left[I\left(\varepsilon_i\le 0\right)\right]+2\tau E\left[I\left(\varepsilon_i>0\right)\right]\\
&=2(1-\tau)F(0)+2\tau\left(1-F(0)\right)=2g(\tau).
\end{aligned}
$$

Therefore, when $n\to\infty$, we have

$$
\begin{aligned}
\mathrm{E}\left(\tilde{G}_{1n}(\gamma)\right)&=\sum_{i=1}^{n}M\left(\frac{\frac{\partial f\left(X_i,\beta\right)}{\partial\beta}\gamma}{\sqrt{n}}\right)\\
&=\sum_{i=1}^{n}\left[g(\tau)\left(\frac{\frac{\partial f\left(X_i,\beta\right)}{\partial\beta}\gamma}{\sqrt{n}}\right)^2+o\left(\left(\frac{\frac{\partial f\left(X_i,\beta\right)}{\partial\beta}\gamma}{\sqrt{n}}\right)^2\right)\right]\\
&=g(\tau)\frac{1}{n}\gamma^T\sum_{i=1}^{n}\frac{\partial f\left(X_i,\beta\right)}{\partial\beta}\frac{\partial^T f\left(X_i,\beta\right)}{\partial\beta}\gamma+o_p(1).
\end{aligned}
$$

According to the A4, we have

$$
E\left(G_{1n}(\gamma)\right)=g(\tau)\gamma^T\Sigma\gamma+o_p(1). \tag{7}
$$

Now that $m_i\left(\hat{\theta}_n\right)-m_i(\theta)=\nabla^T m_i(\theta)\left(\hat{\theta}_n-\theta\right)(1+o(1))$ and Remark 1, we get

$$
\max_i\left|m_i\left(\hat{\theta}_n\right)-m_i(\theta)\right|=o_p(1).
$$

Similarly to the calculation of $G_{1n}(\gamma)$, we can verify $E\left(G_{2n}(\gamma)\right)$,

$$
\begin{aligned}
&E\left(G_{2n}(\gamma)\right)\\
&=E\left(\sum_{i=1}^{n}\frac{1}{1-G_n\left(Y_i\right)}\left(\rho_\tau\left(\varepsilon_i-\frac{\frac{\partial f\left(X_i,\beta\right)}{\partial\beta}\gamma}{\sqrt{n}}\right)-\rho_\tau\left(\varepsilon_i\right)\right)\right) \qquad (8)\\
&=o_p(1).
\end{aligned}
$$

Now, we calculate $E\left(G_{3n}(\gamma)\right)$, From Lemma 1 has $\hat{G}_i-G_i=O_p\left(n^{-1/2}\right)$, we can get $\max_i\left|\hat{G}_i-G_i\right|=O_p(1)$, for large n,

$$
\begin{aligned}
E\left(G_{3n}(\gamma)\right)&\le E\left\{C\max_i\left|\hat{G}_i-G_i\right|\sum_{i=1}^{n}\frac{1}{1-G\left(Y_i\right)}*\right.\\
&\left.\left(\rho_\tau\left(\varepsilon_i-\frac{\frac{\partial f\left(X_i,\beta\right)}{\partial\beta}\gamma}{\sqrt{n}}\right)-\rho_\tau\left(\varepsilon_i\right)\right)\right\} \qquad (9)\\
&=o_p(1).
\end{aligned}
$$

From (7), (8), and (9), it follows that

$$
E\left(G_n(\gamma)\right)\triangleq g(\tau)\gamma^T\Sigma\gamma+o_p(1).
$$

Then, when $\eta_i\in\left(\varepsilon_i-\frac{\frac{\partial f\left(X_i,\beta\right)}{\partial\beta}\gamma}{\sqrt{n}},\varepsilon_i\right)$, at ε_i we can by Taylor expansion of function $\rho_\tau\left(\varepsilon_i-\frac{\frac{\partial f\left(X_i,\beta\right)}{\partial\beta}\gamma}{\sqrt{n}}\right)$, we have

$$\rho_\tau\left(\varepsilon_i-\frac{\frac{\partial f(X_i,\beta)}{\partial\beta}\gamma}{\sqrt{n}}\right)$$

$$=\rho_\tau(\varepsilon_i)-\nabla_{\varepsilon_i}\rho_\tau(\varepsilon_i)\left(\frac{\frac{\partial f(X_i,\beta)}{\partial\beta}\gamma}{\sqrt{n}}\right)$$

$$+\frac{1}{2n}\nabla^2_{\varepsilon_i}\rho_\tau(\eta_i)\gamma^T\frac{\partial f(X_i,\beta)}{\partial\beta}\frac{\partial^T f(X_i,\beta)}{\partial\beta}\gamma+o_p(1),$$

$$\rho_\tau\left(\varepsilon_i-\frac{\frac{\partial f(X_i,\beta)}{\partial\beta}\gamma}{\sqrt{n}}\right)-\rho_\tau(\varepsilon_i)+\rho'_\tau(\varepsilon_i)\left(\frac{\frac{\partial f(X_i,\beta)}{\partial\beta}\gamma}{\sqrt{n}}\right)$$

$$=\frac{1}{2n}\nabla^2_{\varepsilon_i}\rho_\tau(\eta_i)\gamma^T\frac{\partial f(X_i,\beta)}{\partial\beta}\frac{\partial^T f(X_i,\beta)}{\partial\beta}\gamma+o_p(1).$$

In fact, $\left|\nabla^2_{\varepsilon_i}\rho_\tau(\eta_i)\right|\le 2\max(\tau,1-\tau)$ based on $b>0$, s.t., we can

$$\left|L_{in}(\gamma)\right|$$

$$=\left|\frac{\hat{\delta}_i^{ca}}{1-\hat{G}_n(Y_i)}*\right.$$

$$\left.\left[\rho_\tau\left(\varepsilon_i-\frac{\frac{\partial f(X_i,\beta)}{\partial\beta}\gamma}{\sqrt{n}}\right)-\rho_\tau(\varepsilon_i)+\frac{\frac{\partial f(X_i,\beta)}{\partial\beta}\gamma}{\sqrt{n}}\rho'_\tau(\varepsilon_i)\right]\right|$$

$$\le b\max(\tau,1-\tau)\frac{1}{\sqrt{n}}\gamma^T\frac{\partial f(X_i,\beta)}{\partial\beta}\frac{\partial^T f(X_i,\beta)}{\partial\beta}\gamma.$$

Then there exists a constant $c=c(\tau)>0$, s.t., we can

$$\sum_{i=1}^n E\left[\left(L_{in}(\gamma)\right)\right]^2\le c(\tau)\left[\gamma^T\frac{\partial f(X_i,\beta)}{\partial\beta}\frac{\partial^T f(X_i,\beta)}{\partial\beta}\gamma\right]*$$

$$\max_{1\le i\le n}\left(\frac{\left\|\frac{\partial f(X_i,\beta)}{\partial\beta}\right\|^2}{n}\|\gamma\|^2\right)\to 0,$$

where $\|\cdot\|$ denotes the Euclidean norm operator. Here is the final step of convergence to zero, because

$$\gamma^T\frac{\partial f(X_i,\beta)}{\partial\beta}\frac{\partial^T f(X_i,\beta)}{\partial\beta}\gamma\to\gamma^T\Sigma\gamma\ ,$$

$\max_{1\le i\le n}\frac{\left\|\frac{\partial f(X_i,\beta)}{\partial\beta}\right\|^2}{n}\to 0$, for fixed γ, due to the cancellation of cross-product terms, we obtain

$$E\left[\sum_{i=1}^n L_{in}(\gamma)-E\left(\sum_{i=1}^n L_{in}(\gamma)\right)\right]^2$$

$$=\sum_{i=1}^n E\left[L_{in}(\gamma)-E\left(L_{in}(\gamma)\right)\right]^2\le\sum_{i=1}^n E\left[\left(L_{in}(\gamma)\right)\right]^2\to 0.$$

It can be obtained from the above formula, we can

$$\sum_{i=1}^n L_{in}(\gamma)-E\left(\sum_{i=1}^n L_{in}(\gamma)\right)=o_p(1). \quad (10)$$

From (6) to (10), we can obtain

$$G_n(\gamma)\to G_0(\gamma)=g(\tau)\gamma^T\Sigma\gamma+W^T\gamma+o_p(1). \quad (11)$$

The convexity of the limit objective function $G_0(\gamma)$ guarantees the only existence of minimization, therefore we have

$$\sqrt{n}\left(\hat{\beta}-\beta_0\right)=\hat{\gamma}=Arg\min G_n(\gamma)\to\hat{\gamma}_0=Arg\min G_0(\gamma).$$

Derivative for (11): $W^T+g(\tau)\Sigma\gamma=0$, it yields that $\hat{\gamma}_0=g^{-1}(\tau)\Sigma^{-1}W^T$, we have

$$\sqrt{n}\left(\hat{\beta}-\beta_0\right)\xrightarrow{D}N\left(0,\frac{1}{4g^2(\tau)}\Sigma^{-1}\Omega_1\Sigma^{-1}\right).$$

The same can be said

$$\sqrt{n}\left(\hat{\beta}^I-\beta_0\right)\xrightarrow{D}N\left(0,\frac{1}{4g^2(\tau)}\Sigma^{-1}\left(\Omega_1+\Omega_2\right)\Sigma^{-1}\right).$$

IV. Simulation Study

In this section, we illustrate several simulation studies to evaluate the finite sample properties of the proposed estimation methods.

We assume the following non-linear model

$$T_i=e^{\beta_1X_{i1}+\beta_2X_{i2}}+\varepsilon_i, i=1,2,3,\cdots,n$$

where $\beta_1=2,\beta_2=1$, $X_{i1}\sim U(1,2)$, $X_{i2}\sim U(0,0.5)$ are i.i.d., and the model error ε_i are i.i.d.. Let the censoring time C_i is from $\exp(\mu)$, $Y_i=\min(T_i,C_i)$, and $\delta_i=I(T_i\le C_i)$, where μ is adjusted based on different censoring ratio (CR). For the missing mechanism and δ_i is MAR, we assume

$$P\left(\xi_i\left|X_{i1},X_{i2},Y_i\right.\right)=1\big/\left[1+\exp\left(-\alpha_1-\alpha_2X_{i1}-\alpha_3X_{i2}-\alpha_4Y_i\right)\right],$$

where $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$ is adjusted based on different missing ratio (MR). We suppose that $m_0(Y_i,X_{i1},X_{i2},\theta)$ follows a logistic model, that is

$$\log it\left(m_0\left(Y_i,X_{i1},X_{i2},\theta\right)\right)=\theta_1+\theta_2X_{i1}+\theta_3X_{i2}+\theta_4Y_i.$$

The sample size is set to be $n=300$ and 500, the number of simulations is $M=200$, and choose kernel function $K(x)$ to be the Epanechnikov kernel $K(x)=3/4\left(1-x^2\right)_+$. In addition, in this simulation, we take $CR=10\%,30\%$, $MR=10\%,30\%$, respectively. These are achieved by adjusting different μ, $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$ and $\boldsymbol{\theta}=(\theta_1,\theta_2,\theta_3,\theta_4)$. Moreover, we need to consider the following different distributions of the random error: (1) standard normal distribution: $\varepsilon_i\sim N(0,1)$; (2) t-distribution: $\varepsilon_i\sim t(2)$. We are considering the following approaches: the expectile regression estimator given that regression calibration approach (C-ER) and imputation approach (I-ER), for each setting, bias (Bias), mean square error (MSE) and standard deviation (SD) of the C-ER ($\tau=0.25$, 0.5 and 0.75), and I-ER ($\tau=0.25$, 0.5 and 0.75) estimators of parameter β_1,β_0 are summarized in Table I-IV.

From Table I-IV in Appendix, it can be seen that: (i) for each setting, as the sample size n increases, the performance of estimators from both the calibration and the imputation methods improve; (ii) as the censoring ratio (CR) and missing ratio (MR) increase, the estimation performance deteriorates; (iii) regardless of whether ε follows a standard normal distribution or a t-distribution, the ER ($\tau = 0.5$) estimators outperform both ER ($\tau = 0.25$) and ER ($\tau = 0.75$); (iv) for any given censoring ratio and missing ratio, the standard deviation (SD) and mean square error (MSE) of the simulation results obtained from the calibration method are smaller than those from the imputation method, indicating that the calibration method performs better than the imputation method.

Table I
THE SIMULATION RESULTS OF THE BIAS, MES AND SD FOR $n = 300, \varepsilon_i \sim N(0,1)$

CR	MR	Methods	β_1			β_2		
			Bias	MSE	SD	Bias	MSE	SD
10%	10%	C-ER (0.5)	0.35461	0.00206	0.44377	1.91732	0.06070	2.46374
		C-ER (0.25)	1.22981	0.01740	0.55978	2.48233	0.10962	3.05053
		C-ER (0.75)	0.98779	0.01243	0.53804	2.44955	0.09659	2.86376
		I-LS (0.5)	0.35431	0.00206	0.44716	1.96761	0.06316	2.51305
		I-ER (0.25)	1.20829	0.01742	0.57046	2.59956	0.11855	3.20105
		I-ER (0.75)	1.00007	0.01268	0.53898	2.45610	0.09756	2.88261
	30%	C-ER (0.5)	0.36268	0.00213	0.45290	1.93355	0.06268	2.50351
		C-ER (0.25)	1.23213	0.01764	0.57424	2.55272	0.11365	3.12215
		C-ER (0.75)	0.99126	0.01253	0.54409	2.46791	0.09827	2.89234
		I-LS (0.5)	0.36099	0.00214	0.45415	1.96080	0.06398	2.52933
		I-ER (0.25)	1.21729	0.01804	0.57928	2.60140	0.11924	3.20495
		I-ER (0.75)	1.00090	0.01271	0.54510	2.47446	0.09843	2.90119
30%	30%	C-ER (0.5)	0.49802	0.00394	0.61107	2.49253	0.10399	3.22175
		C-ER (0.25)	1.48776	0.02613	0.74168	3.08984	0.16060	3.76411
		C-ER (0.75)	1.12160	0.01759	0.74313	3.09887	0.15876	3.84162
		I-LS (0.5)	0.49990	0.0040	0.62056	2.62846	0.11170	3.34076
		I-ER (0.25)	1.45142	0.02669	0.78053	3.40547	0.18618	4.11106
		I-ER (0.75)	1.12981	0.01781	0.74411	3.15799	0.16245	3.87961

Table II
THE SIMULATION RESULTS OF THE BIAS, MES AND SD FOR $n = 500, \varepsilon_i \sim N(0,1)$

CR	MR	Methods	β_1			β_2		
			Bias	MSE	SD	Bias	MSE	SD
10%	10%	C-ER (0.5)	0.25034	0.00104	0.31660	1.37648	0.03073	1.73258
		C-ER (0.25)	1.18034	0.01546	0.41765	2.21295	0.07125	2.16488
		C-ER (0.75)	1.00619	0.01160	0.38398	1.88369	0.05212	2.07229
		I-LS (0.5)	0.26103	0.00111	0.33009	1.40656	0.03209	1.77007
		I-ER (0.25)	1.15807	0.01568	0.45229	2.28177	0.07762	2.31001
		I-ER (0.75)	1.01822	0.01189	0.38963	1.87334	0.05217	2.07478
30%	30%	C-ER (0.5)	0.38467	0.00230	0.45379	1.83025	0.05524	2.34970
		C-ER (0.25)	1.44795	0.02360	0.55077	2.40331	0.09185	2.72849
		C-ER (0.75)	1.07356	0.01444	0.56171	2.35991	0.09411	2.81729
		I-LS (0.5)	0.39392	0.00243	0.47584	1.90292	0.06106	2.47022
		I-ER (0.25)	1.41137	0.02397	0.61261	2.61716	0.10998	3.04604
		I-ER (0.75)	1.09101	0.01487	0.57163	2.38524	0.09727	2.88018

Table III
THE SIMULATION RESULTS OF THE BIAS, MES AND SD FOR $n = 300, \varepsilon_i \sim t(2)$

CR	MR	Methods	β_1			β_2		
			Bias	MSE	SD	Bias	MSE	SD
10%	10%	C-ER (0.5)	0.86566	0.01303	1.03054	3.96860	0.26850	5.15399
		C-ER (0.25)	2.89352	0.11304	1.73155	5.55598	0.57960	7.09602
		C-ER (0.75)	1.61650	0.03982	1.27423	5.42255	0.49081	6.85035
		I-LS (0.5)	1.04890	0.02244	1.42198	5.68773	0.61968	7.70326
		I-ER (0.25)	2.92856	0.15714	2.85048	9.71510	2.18368	14.39755
		I-ER (0.75)	1.76051	0.04681	1.43964	6.14784	0.62005	7.83334
	30%	C-ER (0.5)	0.93149	0.01510	1.12205	4.01988	0.27972	5.26017
		C-ER (0.25)	2.93450	0.12379	1.96818	5.76487	0.63475	7.47163
		C-ER (0.75)	1.63282	0.04107	1.31810	5.45363	0.49182	6.86661
		I-LS (0.5)	1.01228	0.02339	1.43895	5.34145	0.63865	7.96403
		I-ER (0.25)	2.91836	0.16069	2.90262	9.14646	2.21025	14.55924
		I-ER (0.75)	1.74401	0.04721	1.44260	5.93292	0.65274	7.96468
30%	30%	C-ER (0.5)	1.25761	0.02483	1.33665	4.87671	0.39908	6.18785
		C-ER (0.25)	3.64349	0.17331	2.03036	6.40179	0.73776	7.70207
		C-ER (0.75)	1.70217	0.05020	1.65324	6.38323	0.67366	8.16364
		I-LS (0.5)	1.42685	0.03642	1.81054	7.10242	0.71878	9.33903
		I-ER (0.25)	3.64504	0.22919	3.39743	12.35430	3.55763	18.2304
		I-ER (0.75)	1.87739	0.06148	1.84859	7.18640	0.88877	9.41371

Table IV
THE SIMULATION RESULTS OF THE BIAS, MES AND SD FOR $n = 500, \varepsilon_i \sim t(2)$

CR	MR	Methods	β_1			β_2		
			Bias	MSE	SD	Bias	MSE	SD
10%	10%	C-ER (0.5)	0.82372	0.01055	0.78902	3.21830	0.16520	4.06423
		C-ER (0.25)	3.10769	0.11305	1.30671	4.52452	0.31932	5.28188
		C-ER (0.75)	1.36178	0.02658	0.94792	4.61890	0.34429	5.47718
		I-LS (0.5)	0.82668	0.01129	0.95090	4.14842	0.31157	5.56039
		I-ER (0.25)	2.89602	0.11354	1.83333	7.45612	1.11786	10.14173
		I-ER (0.75)	1.46737	0.03030	0.97872	4.84668	0.37389	5.83341
30%	30%	C-ER (0.5)	1.13968	0.02013	1.08473	4.03713	0.26611	5.15294
		C-ER (0.25)	3.61670	0.15870	1.67031	5.17811	0.44314	6.17712
		C-ER (0.75)	1.43543	0.03356	1.28528	5.53664	0.52632	6.96675
		I-LS (0.5)	1.18731	0.02418	1.33697	5.09346	0.49042	6.99555
		I-ER (0.25)	3.61364	0.18818	2.47862	8.55701	1.55426	12.17887
		I-ER (0.75)	1.52641	0.03741	1.35651	5.98445	0.58556	7.39003

V. A REAL DATA EXAMPLE

In this section, we illustrate the practical applications of the proposed estimation methods through an analysis of datasets in the R software. The GBSG dataset contains patient records from a 1984-1989 trial conducted by the German Breast Cancer Study Group (GBSG) of 686 patients with node positive breast cancer [22]. This dataset comprises 686 observations and 11 variables: recurrence-free survival time (rfstime), age, menopausal status (meno), size, grade, number of positive lymph nodes (nodes), progesterone receptors (pgr), estrogen receptors (er), hormonal therapy (hormon), and event status. The following non-linear model is employed in this paper to fit the data:

$$Y = e^{\beta_1 X_1 + \beta_2 X_2} + \varepsilon \tag{12}$$

where Y, X_1, and X_2 correspond to recurrence free survival time (rfstime), size, and number of positive lymph nodes (nodes), respectively.

From the censoring indicators in the dataset, 299 data sets are censored, with a censoring ratio of 43.586%. To study the expectile regression of the non-linear model under random missingness of censoring indicators, the data missing ratio is artificially defined. Using the simulation method in Section 4, let

$P(\xi_i | X_1, X_2, Y) = 1 / [1 + \exp(-\alpha_1 - \alpha_2 X_1 - \alpha_3 X_2 - \alpha_4 Y)]$, we can obtain

$P\left(\xi_i \middle| X_1, X_2, Y\right) = 1 \big/ \left[1 + \exp\left(0.15 - 0.01X_1 + 1.25X_2 - 0.1745Y\right)\right]$, and a missing ratio of 10.7%. After introducing missing data, among 686 patients, 344 were uncensored with observed censoring indicators ($\delta = 1, \xi = 1$), 263 were censored with observed censoring indicators ($\delta = 0, \xi = 1$), and 79 had randomly missing censored indicators ($\xi = 0$). To compare the effects of different estimation methods more intuitively and effectively, we define the overall mean square error (MSE) as $AMSE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2$, with results present in Table V.

From Table V in Appendix, the following observations can be made: (i) with a censoring ratio of 43.586% and a missing ratio of 10.7%, the ER ($\tau = 0.5$) estimators outperform both ER ($\tau = 0.25$) and ER ($\tau = 0.75$); (ii) the mean square error (MSE) of the results obtained via the calibration method is smaller than that of the imputation method, indicating that the calibration method performs better than the imputation method.

Table V
THE AMSE UNDER DIFFERENT ESTIMATION METHODS

	method	MSE
$\tau = 0.25$	C-ER	0.9381375
	I-ER	0.9584808
$\tau = 0.5$	C-ER	0.7548942
	I-ER	0.7574275
$\tau = 0.75$	C-ER	0.7970802
	I-ER	0.7985787

REFERENCES

[1] W. Newey, J. L. Powell, “Asymmetric least squares estimation and testing,” *Econometrica*, vol. 55, no. 4, pp. 819-847, 1987.

[2] L. S. Waltrup, F. Sobotka, T. Kneib, G. Kauermann, “Expectile and quantile regression—David and Goliath?,” *Statistical Modelling*, vol. 15, no. 5, pp. 433 - 456, 2015.

[3] F. Sobotka, G. Kauermann, L. S. Waltrup, T. Kneib, “On confidence intervals for semiparametric expectile regression,” *Statistics and Computing*, vol. 23, pp. 135-148, 2013.

[4] M. Kim, S. Lee, “Nonlinear expectile regression with application to Value-at-Risk and expected shortfall estimation,” *Computational Statistics and Data Analysis*, vol. 94, pp. 1-19, 2016.

[5] R. Jiang, Y. X. Peng, Y. F. Deng, “Variable selection and debiased estimation for single - index expectile model,” *Australian and New Zealand Journal of Statistics*, vol. 63, no. 4, pp. 658-673, 2022.

[6] S. H. Gao, Z. Yu, “Parametric expectile regression and its application for premium calculation,” *Insurance: Mathematics and Economics*, vol. 111, pp. 242-256, 2023.

[7] O. Litimein, A. Laksaci, B. Mechab, S. Bouzebda, “Local linear estimate of the functional expectile regression,” *Statistics and Probability Letters*, vol. 192, pp. 109682, 2023.

[8] X. B. Ji, S. H. Luo, M. J. Liang, “Quantile regression for single-index varying-coefficient models with missing movariates at random,” *IAENG International Journal of Applied Mathematics*, vol. 54, no. 6, pp. 1117-1124, 2024.

[9] H. L. Koul, V. Susarla, J. V. Ryzin, “Regression Analysis with Randomly Right-Censored Data,” *Annals of Statistics*, vol. 9, no. 6, pp. 1276-1288, 1981.

[10] Y. L. Pan, Z. Liu, G. Y. Song, “Weighted expectile regression with covariates missing at random,” *Communications in Statistics - Simulation and Computation*, vol. 52, no. 3, pp. 1057-1076, 2021.

[11] A. Seipp, V. Uslar, D. Weyhe, A. Timmer, F. Otto-Sobotka, “Weighted expectile regression for right-censored data,” *Statistics in Medicine*, vol. 40, no. 25, pp. 5501-5520, 2021.

[12] Q. Zhao, Z. D. Wang, J. J. Wu, X. L. Wang, “Weighted expectile average estimation based on CBPS with responses missing at random,” *AIMS Mathematics*, vol. 9, no. 8, pp. 23088-23099, 2024.

[13] F. P. Zhang, X. Chen, P. Liu, C. Y. Fan, “Weighted expectile regression neural networks for right censored data,” *Statistics in Medicine*, vol. 43, no. 27, pp. 5100-5114, 2024.

[14] G. Ciuperca, “Right-censored models by the expectile method,” *Lifetime Data Analysis*, vol. 31, no. 1, pp. 149-186, 2025.

[15] Q. H. Wang, G. E. DINSE, “Linear regression analysis of survival data with missing censoring indicators,” *Lifetime Data Analysis*, vol. 17, no. 2, pp. 256-279, 2010.

[16] Y. Shen, H. Y. Liang, “Quantile regression for partially linear varying-coefficient model with censoring indicators missing at random,” *Computational Statistics and Data Analysis*, vol. 117, pp. 1-18, 2018.

[17] J. F. Wang, W. J. Jiang, F. Y. Xu, W. X. Fu, “Weighted composite quantile regression with censoring indicators missing at random,” *Communications in Statistics - Theory and Methods*, vol. 50, no. 12, pp. 2900-2917, 2019.

[18] X. S. Zhou, P. X. Zhao, “Estimation and inferences for varying coefficient partially nonlinear quantile models with censoring indicators missing at random,” *Computational Statistics*, vol. 37, no. 4, pp. 1727-1750, 2022.

[19] Z. W. Cai, “Asymptotic properties of Kaplan-Meier estimator for censored dependent data,” *Statistics and Probability Letters*, vol. 37, no. 4, pp. 381-389, 1998.

[20] X. Y. Li, Q. H. Wang, “The weighted least square based estimators with censoring indicators missing at random,” *Journal of Statistical Planning and Inference*, vol. 142, no. 11, pp. 2913-2925, 2012.

[21] Q. H. Wang, K. W. NG, “Asymptotically efficient product-limit estimators with censoring indicators missing at random,” *Statistica Sinica*, vol. 18, no. 2, pp. 749-768, 2008.

[22] W. Sauerbrei, P. Royston, H. Bojar, C. Schmoor, M. Schumacher, “Modelling the effects of standard prognostic factors in node-positive breast cancer. German Breast Cancer Study Group (GBSG),” *British Journal of Cancer*, vol. 79, no. 11-12, pp. 1752-1760, 1999.